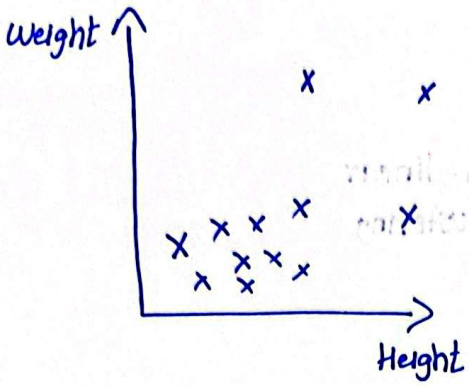
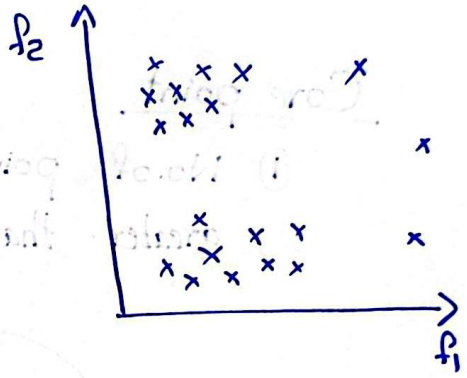
Anomaly Detection

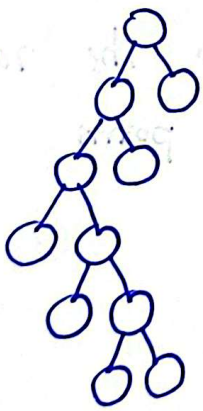


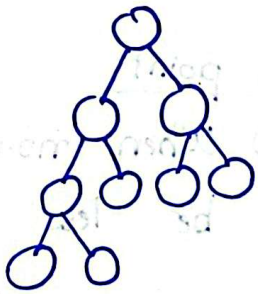
⟶ Isolation Forest

f_1	f_2	f_3	f_4
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–



Isolation forest try to create a leaf node for every data point.





$E(h(x) > c(m)$
$\Rightarrow S(x,m) \approx 0.5 \Rightarrow$ Normal data point

$E(h(x) < c(m) \Rightarrow S(x,m) \approx 1$
⇓
anomaly score is good/outlier

Formula

Anomaly score of a new point

$$S(x,m) = 2^{\frac{-E(h(x))}{c(m)}}$$

data point ← $S(x,m)$ → no of data point

$h(x)$ = average search depth for x from isolated tree

$c(m)$ = average value of $h(x)$

⟶ DBSCAN Clustering

① Core Point
② Border Point
③ Noise/outlier

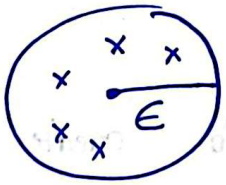
} Non-linear clustering

hyperparameters

① minimum points
② epsilon => radius

Core point

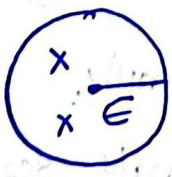
① No. of points within the radius should be greater than minimum points



minimum point = 4

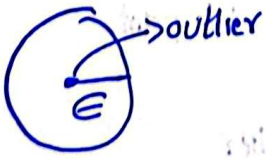
Border point

① When no.of points within the radius will be less than minimum points

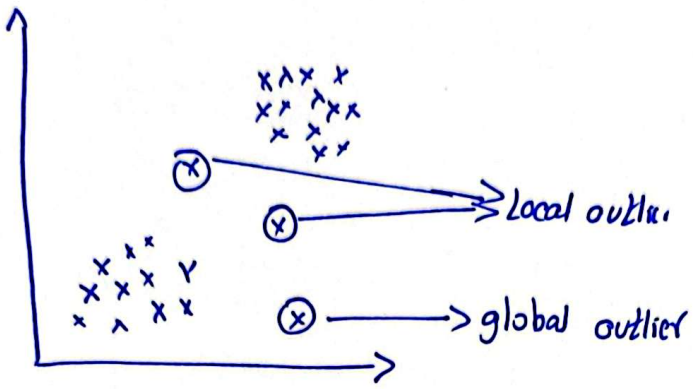


Outliers

DBSCAN is Robust to outliers



→ Local Outlier Factor



① Local outlier

② Global outlier

It uses K-nearest neighbour concept

↳ try to determine local density

Steps:

① Pick the number of K

② Find nearby points for each point

③ Calculate distance

④ Measure local density

⑤ Compare densities

a. If density of points is lower than its neighbor's density, it is likely an outlier

⑥ Identify outliers